Columbia University Automatic Video Search in TRECVID 2006

Shih-Fu Chang, Winston Hsu, Wei Jiang, Lyndon Kennedy, Dong Xu, Akira Yanagawa, and Eric Zavesky

Objectives and Unique Features

+ Objective: Enable semantic search over a large video collection. Leverage visual content and text.
  + Text-based queries against visual content.
  + Concept detectors predict the concept in image.
  + Lexicon of definitions maps keyword to concept.

Search System Overview

Concept-Based Search

+ 374 detectable concepts, compared to standard 39.
+ 10x increase in lexicon ~ 30% increase in search

Large-Scale Concept Lexicon

+ Leverage LSCOM annotations. (449 concepts).
+ 374 detectable concepts, compared to standard 39.
+ 10x increase in lexicon ~ 30% increase in search

Search System Overview

Multimodal Query

Text: Find tall buildings.
Visual Examples

Concept-Based Search

Text: Find soccer.
Visual Examples

Fused Search Result

Named Person: if named entity detected in query. Rely on text search.
Sports: if sports keyword detected in query. Rely on visual examples.
Concept: if keyword maps to concept detector. Rely on concept search.
Named Person + Concept: if both named entity and concept detected. Combine text and concept search equally.

Class Preparation

+ Select classes to suit tools.
+ Weights optimize AP (over training queries).
+ Classify by query keywords. (automatically at query time)

Class/Coverage

Named Person
Sports
Concept
Name + Concept
General

Evaluation

+ Tested over 160+ hours of multilingual news video: TRECVID 2006.
+ Consistent gains in P@100 (Precision of top 100) through class-dependent multimodal fusion.
+ Text and Concept-based searches perform well.
+ Most query topics have matching concepts.